A NOVEL FRAMEWORK FOR SMART AGRICULTURE PREDICTION USING MACHINE LEARNING TECHNIQUES

Introduction:

State of the Art:

Agriculture plays a significant role in the financial system worldwide, demand in the agricultural system will rise with the ongoing growth of the human population. Agriculture - technology and precision farming, now also termed digital agriculture, have arisen as a novel technical field that use data centric approaches to have better yield. The data created in current agricultural operations given by a variety of sensors that allow a superior understanding of the operational environment (an interaction of dynamic soil, crop, and weather circumstances) and the operation itself (machinery data), leading to extra precise and quicker decision making. This research primarily aims to enhance agriculture technology. The research methodology is categorized into four phases, viz soil profile prediction, crop yield production forecasting, crop disease detection, groundwater and dam reservoir level forecasting. The experimental result shows that the proposed Modified Regression, Fuzzy C Means, Linear Discriminant Analysis and Artificial Neural Network model gives superior performance than the existing models.

Motivation:

Soil quality is degraded based on soil acidification and alkalization. These processes affect agricultural production and water quality. Accurate Soil profile prediction is required to solve this problem before planting a crop to assure that it will be getting the right amount of nutrients from the soil.

Forecasting crop yield production has been a constant concern since the beginning of the history of agriculture. Many forecasting techniques have evolved in the agriculture. Those who use forecast data are always seeking greater accuracy, granularity, comparability, and timeliness.

Crop production forecasting is central to making food policy decisions in developing countries. Almost all major food security programs—such as food aid imports, strategic food reserves, granting of licenses for private firms to import or export, local procurement by the

government and donors, emergency food assistance, and distribution through social safety net programs—rely on crop forecasts for strategic planning.

Pests and Diseases destroy crops results in decreased food production, leading to food insecurity. Also, knowledge about pest management or control and diseases are less in various underdeveloped countries. Toxic pathogens, poor disease control, drastic climate changes are one of the critical factors which arise in dwindled food production. Different modern technologies have emerged to minimize post-harvest processing, to fortify agricultural sustainability and to maximize productivity. For disease identification and prediction, Polymerase chain reaction, gas chromatography and mass spectrometry techniques employed. However, these techniques are not cost-effective and are highly time-consuming. It motivates to crop disease detection and prediction.

Using groundwater for agricultural production has the potential to build resilience in food-insecure regions of the world. Use of groundwater can boost agricultural production, improve rural revenue and strengthen farmers' ability to withstand climate shocks and water variability. Since groundwater level varies in a different land, it paves way for predicting groundwater level.

The reservoir is a physical structure such as a pond or lake, either natural or artificially developed to impound and regulate the water. It has used as one of the structural approaches to enhance agricultural, flood defence and water storage. During both flood and drought situations, the decision to open or close watergate is a critical action that needs to be undertaken by dam operator as a late decision will not only cause flood downstream but also will damage dam structure.

Releasing the water earlier before the reservoir reaching its full capacity might reduce the flood risk downstream. However, one cannot be sure that released water will be stored and used when there is less rainfall. If the dam water is used for multipurpose, low water in the reservoir will cause conflict on its usage. Researchers believe that the use of forecasting and warning system might improve the dam operation and decision. It motivates dam reservoir level prediction.

Problem Definition:

Over the last decades, many problems occur in the agricultural work especially in predicting Soil Profile (acid and alkaline), crop yield, crop disease, groundwater level decreasing and wrong decision of dam reservoir level,etc. Based on these, this research has attempted to find solutions for the same. The issues taken are as follows:

How to predict the soil profile? How and what method used to predict the soil profile?

How to estimate crop yield production?

How to detect and predict crop disease?

How to monitor and predict groundwater level and dam reservoir level?

Objectives of the Research:

- **Primary Objective**
  - To propose Soil Profile Prediction through Modified Regression by Discretization algorithm for the Crop yield in Trichy district.
- **Secondary Objectives**
  - To predict Crop Yield in Trichy District using Fuzzy C Means Algorithm and Multilayer Perceptron (MLP).
  - To develop, an efficient, Crop disease prediction mechanism for the Indian Banana, Paddy and Sugarcane using the Linear Discriminant Analysis (LDA) algorithm.
  - To propose a framework for forecasting groundwater and dam reservoir level based on an artificial neural network.

Review of Literature:

John Carlo Puno et al. [1] discussed the soil nutrients and pH level using Image Processing and Artificial Neural Network to efficiently identify the nutrients and pH level of soil with the use of Soil Test Kit (STK) and Rapid Soil Testing (RST) of the Bureau of Soils and Water Management. They used Artificial Neural Network to hasten the performance of image processing in giving an accurate result. Amati et al. [2] proposed a retrieval algorithm for

surface-to-root-zone profiles of soil moisture (RZSM). The algorithm uses reflectometry signals at L-band, P-band, and their combinations to obtain accurate estimates of RZSM. First, they use an RZSM profile scattering model that contains both coherent and incoherent scattering contributions. There are multiple options to solve this problem, and their group has previously developed both numerical and approximate analytical methods for this purpose. Second, they make a correspondence between polarimetric scattering cross-sections, and, circularly polarised Delay Doppler Map (DDM) observations. The next step is to solve the inverse problem for surface soil moisture and RZSM using single frequencies (L- and P-band) and their combinations if available within a reasonable observation time window.

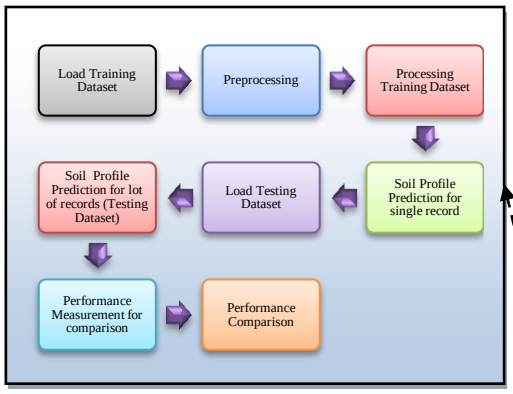
To accomplish the latter task, they use a powerful inversion method recently developed in their group, which is a hybrid of global and local optimization methods. They perform a thorough sensitivity analysis to investigate the utility of either frequency alone and in combination for inverting soil moisture at the surface and the root zone. They then show the application of this method to actual Signals of opportunity data to the extent available. Recommendations made for the combined use of GNSS and MUOS for retrieving RZSM. Bang et al. [3] used three methods to forecast- ARMA (Auto Regressive Moving Average), SARIMA (Seasonal Auto-Regressive Integrated Moving Average) and ARMAX (ARMA with exogenous variables). The performance of the three compared, and the best model is used to predict rainfall and temperature, which are in turn used to predict the crop yield based on a fuzzy logic model. Gandge et al. [4] offer the study about the different data mining techniques used for forecasting the crop yield. The success of any crop yield forecasting system heavily relies on how accurately the features have extracted and how appropriately classifiers have employed. They summarise the results extracted by different algorithms which are being used by different authors for crop yield prediction, with their accuracy and recommendation. Every other field has got some benefit from new technologies as compared to the agricultural field. According to past studies, 42% of agricultural production is in the loss due to plant leaf diseases, Amrita et al. [5] proposed a method for plant leaf disease detection can be applied to detect a disease from the input images to overcome this significant issue. This process involved steps like image preprocessing, image segmentation, feature extraction. Further K Nearest Neighbor (KNN) classification is applied to the outcome of these three stages. Proposed implementation has shown 98.56% of accuracy in predicting plant leaf diseases. It also presents other information regarding a plant leaf disease

that is Affected Area, Disease Name, Total Accuracy, Sensitivity and Elapsed Time. Producing chili is a cumbersome process because of the plant presented to the assaults from entirely different smaller scale creatures, bacterial disorders and bugs—the signs of attack typically distinguished through leaves, peduncle organic product review. Nowadays, without thinking the requirement of plants chemical is applied to the plant. Patil et al. [6] system will guarantee that the chemical only applied when plant tainted with disease. Chili disease images take through image processing technique. The plant disease has been identified through the leaf image and also various data processing techniques is beneficial and cheap for assessment of diseases of plant for assisting the farmers. Kaya et al. [7] proposed Artificial Neural Networks (ANN) and M5tree (M5T) models investigated to enhance the methods to estimate the groundwater level. The accuracy of the ANN and M5T models in groundwater level estimation also investigated, and the results compared with one another. Üneş et al. [8] proposed generalized regression neural network, support vector machines and m5 decision tree models for predicting dam reservoir level.

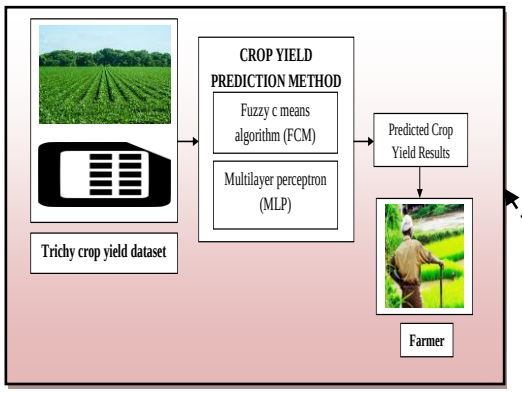
Methodology:

The overall architectural representation of an intelligent framework for smart agriculture prediction using machine learning techniques illustrated in Figure

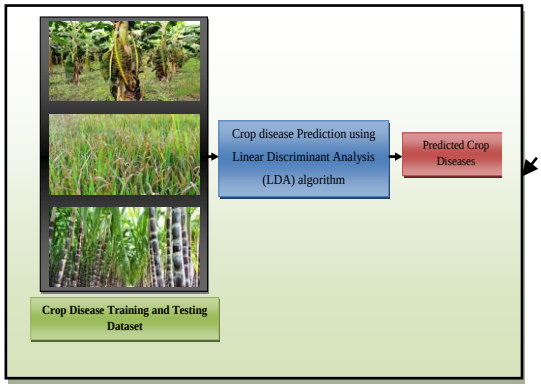
1. This research work is implemented in four Phases.



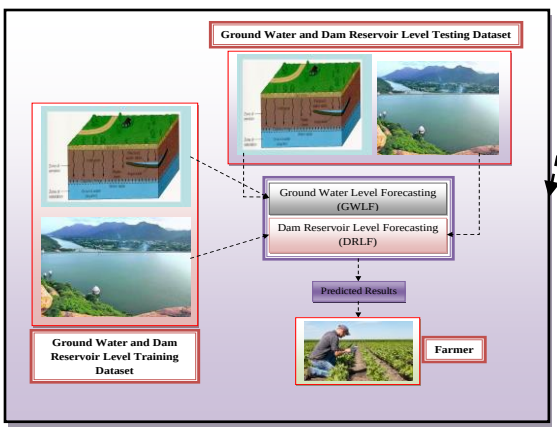
Phase 1



Phase 2


Crop disease Prediction using Linear Discriminant Analysis (LDA) algorithm
Predicted Crop Diseases
Crop Disease Training and Testing Dataset

Phase 3


Ground Water and Dam Reservoir Level Testing Dataset
Ground Water Level Forecasting (GWLF)
Dam Reservoir Level Forecasting (DRLF)
Predicted Results
Farmer
Ground Water and Dam Reservoir Level Training Dataset

Phase 4

A Novel framework for smart agriculture prediction using machine learning techniques

Figure 1: Overall Novel Framework for smart agriculture prediction using machine learning techniques

Phase 1: Soil Profile Prediction through Modified Regression by Discretization algorithm for the Crop yield in Trichy district: This phase has taken up the issue of predicting the yield level by analyzing the soil characteristics for the particular soil. As gathering the information from the enormous volume of data is practically a complicated task in the recent circumstances. Hence, this phase has concentrated only Trichy district, and the soil data samples have been collected. A preprocessing element was premeditated to construct training and testing sets for use among the proposed and compared classifiers. The classifiers like Linear Regression, Simple Linear Regression, Additive Regression, and Regression by Discretization used for training and testing of the data sets. This phase has modified the Regression by Discretization classifier and applied

Result: The modified Regression by Discretization classifier had the least error of the predicted values. The least root relative squared error and the better cover coefficient so henceforth it can be used to forecast soil fertility in advance with greater accuracy when compared to the other classifiers.

Inference: This phase will recommend appropriate fertilizer for the given soil sample and cropping pattern. Though importance is on escalating the current yield level, consideration should also be set to avert soil deprivation and also to widen the appropriate technologies to rescue the problem soils.

Phase 2: Crop Yield Production Forecasting in Trichy District using Fuzzy C Means Algorithm and Multilayer Perceptron: The forecast is made based on the previous archives of yield data secured from that area. The systems examine the data, and on mining, several patterns based on numerous parameters predict the return. This phase directs on crop yield forecast in Trichy district by adopting data mining techniques for rule formation on classifying the training data and implementing prediction for test data. The suggested method employs Fuzzy C means algorithm for clustering and Multilayer Perceptron design for prediction.

Result: The proposed FCM learning approach predicts the phenomenon correctly and gives the right decision about the class of yield.

Inference: The proposed Crop Yield Production Forecasting has proved better performance than the regression models.

Phase 3: Crop Disease Prediction using the Linear Discriminant Analysis (LDA): Data mining approaches such as classification help to detect or predict the crops diseases, production, and loss. The data is collected from the farmers to recognize what type of diseases frequently appear in the crops. Generally, the classification rules and relationships are used for acquiring knowledge from empirical data of diverse dataset. This work explores what Classification rule and prediction can do in the agricultural domain. This work is intended to discover classification rules for the Indian Banana, Paddy and Sugarcane using the Linear Discriminant Analysis (LDA) algorithm.

Result: The proposed LDA method generates more efficient prediction and helps in the design of control systems that minimize yield losses.

Inference: The application of LDA models for crop disease prediction as revealed from the present work on banana, paddy and sugarcane prediction makes them a useful tool for the future forecasting model.

Phase 4: Groundwater and dam reservoir level forecasting using Artificial Neural Network: Ground water is one of the most valuable natural resources for many developing countries. Trichy groundwater level data are taken from 2017 to 2020 for predicting future groundwater level till 2025. Mettur dam reservoir level data are taken from 2017 to 2020 for predicting future reservoir level till 2025. As a sub-approach of ANN, for forecasting groundwater and dam reservoir level, a Multilayer Perceptron (MLP) is used.

Result: The proposed ANN model provides closer accuracy in the groundwater and dam reservoir level, and ANN method can use to predict groundwater level and dam reservoir level efficiently.

Inference: This proposed Ground Water Level Forecasting and Dam Reservoir Level Forecasting algorithm predict groundwater level and dam reservoir level more accurately.

Research Contributions:

This research work aims to develop an intelligent framework for smart agriculture prediction using machine learning techniques, which can provide better prediction results, compared to the native prediction techniques. Thus, to obtain the goal of the research work, following are the contributions of this research work.

1. Developed a soil profile prediction technique through modified regression by discretization algorithm for the crop yield in Trichy district.
2. Proposed a crop yield prediction method in Trichy district using fuzzy c means algorithm and Multilayer Perceptron (MLP).
3. Developed an efficient crop disease prediction mechanism for the Indian banana, paddy and sugarcane using the Linear Discriminant Analysis (LDA) algorithm.
4. Proposed a framework with an artificial neural network proposed for forecasting groundwater and dam reservoir level.

Conclusion:

This research primarily aims to enhance agriculture technology. The research methodology is categorized into four phases. Each layer in the soil has various characteristics and the Soil Profile supports in deciding the role of the soil. It distinguishes the given soil sample from the other soil samples based on its chemical composition. Soil Profile Prediction through Modified Regression by Discretization algorithm for the Crop yield in Trichy district has been done which helps to study the yield level for the various soils in Trichy district, Tamil Nadu using different machine learning algorithms and prediction techniques. Assessment of this results shows that the modified Regression by Discretization classifier had the least error of the predicted values and the least root relative squared error and the better cover coefficient. Henceforth it can be used to forecast soil fertility in advance with better accuracy when compared to the other classifiers. The crop yield forecast helps farmers to decide on crop cultivation based on soil types. Crop Yield Forecasting in Trichy District using Fuzzy C Means Algorithm and Multilayer Perceptron helps to predict the high variability as compared to Regression models and works well when more missing values have occurred in the dataset. The results have also shown that the FCM learning approach predicts the phenomenon correctly and gives the right decision about the class of yield. Some Crop diseases do not have prominence during the early stage, which results in the damage of the whole crop after growing fully. Hence,

Crop Disease Prediction using the Linear Discriminant Analysis (LDA) has been done with an extensive analysis of different data mining classifiers on different feature sets to predict crop diseases efficiently. The higher predictive accuracy by the latest machine learning techniques like LDA, as demonstrated in the present study will generate a more efficient prediction and help in the design of control systems that minimize yield losses. Reservoir can be used to balance the water flow during various weather conditions, such as decreasing the flow by holding water back during heavy rainfall and releasing more water during droughts. Hence, groundwater and dam reservoir level forecasting using Artificial Neural Network is done, which helps to predict the water level well in advance. Comparisons revealed that the ANN model had closer accuracy in the groundwater and dam reservoir level and ANN method can be used to predict groundwater level and dam reservoir level efficiently.

Scope for Further Research:

Despite contributions of this synopsis regarding precision farming, several future directions may be addressed to make further advancement in the area.

- Soil microorganisms can form complex and varied communities which interact with each other in many different ways depending on environmental conditions. Different metabolic paths and adaptability accompany these microbial diversities reflected even in extreme environments.
- In recent decades, the biodiversity of microbes in extreme environments has been in scientific focus because such specifically adapted bacteria can improve bioremediation processes in industrial and agricultural applications. Instead of the time-consuming process of identification of new bacterial strains from habitats rich in the microbiota, artificial neural networks have proposed as a mapping model for resolving the problem of prediction of microbial behaviour.
- The proposed FCM and MLP based crop yield prediction technique cannot handle big dataset. Big data analysis is efficient in handling a large amount of data generated for predicting agricultural yield. In this work, a neural network is used for prediction and to enhance its performance.
- A population-based incremental learning technique (Hadoop) may be proposed for optimizing the weights.

- A new website may be launched or an app may be developed to help farmers for diagnosing crop diseases at an early stage.
- The present study attempted to predict groundwater and dam reservoir levels obtained from previous year data based on ANN may be extended using the radial basis function (RBF) neural network–whale algorithm (WA) model, the multilayer perception (MLP–WA) model, and genetic programming (GP).

LIST OF PUBLICATIONS

CONFERENCE ATTENDED AND PAPER PRESENTED:

1. "Analyzing performance monitoring of different classification techniques for diabetic data" in the 3rd International Conference on science technology and management on 17th Jan 2016 at India International Centre, New Delhi.

2. "Analyzing the suitability of relevant classification on medical data" in the 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) on 10 and 11 Feb 2017 at SCAD Institute of Technology, Palladam, Coimbatore

3. "Applying Data Mining Techniques in Agriculture - A survey" in UGC sponsored National Conference on Digital India - Power to Empower on 22 Sep 2017 at Avinashilingam Institute for Home Science and Higher Education for Women

4. "Applying Machine Learning Techniques in Agriculture to Forecast Crop Yield – A Survey" in International Conference on SMAC on 9th Feb 2018 at Avinashilingam Institute for Home Science and Higher Education for Women.

JOURNALS PUBLISHED:

5. Ms. M.C.S. Geetha & Dr. I. Elizabeth Shanthi, "Predicting the Soil Profile through Modified Regression by Discretization algorithm for the crop yield in Trichy district" in International Journal of Grid and Utility Computing,Vol 9 No 3, pp-235-242,ISSN online: 1741-8488; ISSN print: 1741-847X. On Feb 2018. (scopus & web of science)
6. Ms. M.C.S. Geetha & Dr. I. Elizabeth Shanthi, **"**Forecasting the crop yield production in

Trichy District using Fuzzy C-Means Algorithm and Multilayer Perceptron(MLP)", International Journal of Knowledge and Systems Science,Volume11,Issue3,p-83 to 87,ISSN: 1947-8208, DOI No.10.4018/IJKSS.2020070105 on Sep 2020. (scopus & web of science)

7. Ms. M.C.S. Geetha & Dr. I. Elizabeth Shanthi, **"**Classification and prediction in agricultural domain using the linear discriminant analysis(LDA)", Journal of Green Engineering, Vol 10(12) 2020, PP 13374-13392,ISSN: 1904-4720 (Print) on Dec 2020. (scopus)
8. Ms. M.C.S. Geetha & Dr. I. Elizabeth Shanthi, **"**Forecasting Ground water and Dam Reservoir Level based on Artificial Neural Network", Journal of Green Engineering, Vol 11(1) 2021, PP 179-194,ISSN: 1904-4720 (Print) on Jan 2021. (scopus)
9. Ms. M.C.S. Geetha & Dr. I. Elizabeth Shanthi, "A Survey and Analysis on Regression Data Mining Techniques in Agriculture", International Journal of Pure and Applied Mathematics Volume 118 No. 8,pp- 341-347 ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version) on Jan 2018. (scopus)
10. Ms. M.C.S. Geetha & Dr. I. Elizabeth Shanthi, "Applying Machine Learning Techniques in Agriculture to Forecast Crop Yield – A Survey" in International Journal of Theoretical and Applied Sciences, Vol 10(1), pp-62-67,ISSN NO. Print : 0975-1718, Online: 2249-3247 on Feb 2018. (UGC)
11. Ms. M.C.S. Geetha & Dr. I. Elizabeth Shanthi, "Analyzing the suitability of relevant classification techniques on medical data set for better prediction" 2017 International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC) on Oct 5, 2017. Published in IEEE Xplore (Scopus)
12. Ms. M.C.S. Geetha & Dr. I. Elizabeth Shanthi, "Analyzing performance monitoring of different classification techniques for diabetic data" in the International Journal of advanced technology in engineering and science(IJATES), Volume 4 , Issue 1, Jan 2016, ISSN 2348-7550.
13. Ms. M.C.S. Geetha & Dr. I. Elizabeth Shanthi, "Going Green with IoT for Smart World" in World Scientific News, WSN 41(2016) 167-173.

REFERENCES:

[1] John Carlo, PUNO et al. (2017), "Determination of Soil Nutrients and pH level using Image Processing and Artificial Neural Network", IEEE 9th International Conference on Humanoid, Nanotechnology, Information Technology, Communication and Control, Environment, and Management (HNICEM).

[2] Amir Azemati, Aslan Etminan, Alireza Tabatabaeenejad, Mahta Moghaddam, "Retrieval of Subsurface Soil Moisture Profiles from L-Band and P-band Reflectometry", IEEE International Conference on Electromagnetics in Advanced Applications (ICEAA), Sept. 2019, DOI: 10.1109/ICEAA.2019.8879216.

[3] Shivam Bang, Rajat Bishnoi, Ankit Singh Chauhan, Akshay Kumar Dixit, Indu Chawla, "Fuzzy Logic based Crop Yield Prediction using Temperature and Rainfall parameters predicted through ARMA, SARIMA, and ARMAX models", IEEE Twelfth International Conference on Contemporary Computing (IC3), September 2019, DOI: 10.1109/IC3.2019.8844901.

[4] Yogesh Gandge, Sandhya, "A study on various data mining techniques for crop yield prediction"., Electrical Electronics Communication Computer and Optimization Techniques (ICEECCOT) 2017 International Conference on, pp. 420-423, 2017.

[5] Amrita S. Tulshan, Natasha Raul, "Plant Leaf Disease Detection using Machine Learning", Computing Communication and Networking Technologies (ICCCNT) 2019 10th International Conference on, pp. 1-6, 2019.

[6] Asha Patil, Kalpesh Lad, "Review of Diseases Detection and Classification for Chilli Leaf using Various Algorithms", Electrical Computer and Communication Technologies (ICECCT) 2019 IEEE International Conference on, pp. 1-4, 2019.

[7] Yunus Ziya Kaya, Fatih Üneş, Mustafa Demirci and Bestami Taşar, "Groundwater Level Prediction Using Artificial Neural Network and M5 Tree Models", Air and Water Components of the Environment Conference, March 2018.

[8] Fatih Üneş, Mustafa Demirci, Bestami Taşar, Yunus Ziya Kaya, "Modeling of dam reservoir volume using generalized regression neural network, support vector machines and m5 decision tree models", Applied Ecology and Environmental Research 17(3):7043 • May 2019.